Decomposing the Composition Effect: the Role of Covariates in Determining Between-Group Differences in Economic Outcomes

Christoph Rothe*

Columbia University

Abstract

In this paper, we study the structure of the composition effect, that is the part of the observed between-group difference in the distribution of some economic outcome that can be explained by differences in the distribution of covariates. Using results from copula theory, we derive a new representation that contains three types of components: (i) the "direct contribution" of each covariate due to between-group differences in the respective marginal distributions, (ii) several "two way" and "higher order interaction effects" due to the interplay between two or more marginal distributions, and (iii) a "dependence effect" accounting for between-group differences in dependence patterns among the covariates. We show how these components can be estimated in practice, and use our method to study the evolution of the wage distribution in the US between 1985 and 2005. We obtain some new and interesting empirical findings. For example, our estimates suggest that the dependence effect alone can explain about one fifth of the increase in wage inequality over that period (as measured by the difference between the 90% and the 10% quantile).

*This Version: March 5, 2014. Address for correspondence: Department of Economics, Columbia University, 420 W 118th St, New York, NY 10027, USA. E-Mail: cr2690@columbia.edu. Website: http://www.christophrothe.net.

1. Introduction

Understanding the factors accounting for the differences in the distributions of individuals' economic outcomes across two countries, time periods, or subgroups of the population is central in several fields of economic research, particularly in labor and development economics. For instance, the question why wage inequality has increased substantially in the US and other industrialized countries over the past decades has received enormous attention in the recent literature. Other examples include the study of the gender wage gap, wage differentials between natives and immigrants, and variations in health outcomes across several developing regions. The distributional aspect is often critical in these applications. Comparing real hourly wages among male US workers in 1985 and 2005, for example, one can observe that the median wage has remained approximately constant, but that the 90% and 10% quantile have increased by about 20% and 5%, respectively. There has thus been a substantial change in the overall shape of the wage distribution, which would not be revealed by a simple comparison of means.

With such applications in mind, a number of papers have proposed procedures to decompose between-group differences in economic outcomes into two components: a composition effect due to differences in observable covariates across groups, and a structure effect due to differences in the relationship that links the covariates to the outcome. The most popular example is certainly the Oaxaca-Blinder procedure (Oaxaca, 1973; Blinder, 1973) for decomposing differences in mean outcomes when the data are generated by a simple linear model. More flexible methods that can be used to decompose general distributional features like quantiles or inequality measures, and allow for complex nonlinear relationships between the covariates and the outcome variable are proposed and studied by DiNardo, Fortin, and Lemieux (1996), Gosling, Machin, and Meghir (2000), Donald, Green, and Paarsch (2000), Barsky, Bound, Charles, and Lupton (2002), Machado and Mata (2005), Melly (2005), Rothe (2010), and Chernozhukov, Fernandez-Val, and Melly (2013), among others. See Fortin, Lemieux, and Firpo (2011) for a literature review.

A natural question in empirical applications is whether the composition effect can be further decomposed into contributions attributable to specific covariates. For example, it could be interesting to know what portion of the composition part of the wage gap between natives and immigrants can be attributed to between-group differences in education. Further decompositions of between-group differences in mean outcomes can be obtained via the Oaxaca-Blinder procedure when the data are generated by a linear model. For more general settings, alternative further decompositions have been proposed and studied by e.g. DiNardo et al. (1996), Altonji, Bharadwaj, and Lange (2012), or Chernozhukov et al. (2013), amongst many others. These methods are generally *path dependent*, in the sense that their results depend on the order of the covariates in the data set, which can be undesirable in many applications where there is no natural order among the covariates. More importantly, the elements of these further decompositions typically measure the contribution of between-group differences in the *conditional* distribution of a covariate given the remaining ones, and can thus not be interpreted as reflecting between-group differences in a single, specific covariate.

It turns out that it is not possible to develop a general method that apportions the composition effect into components that are only attributable to between-group differences in the marginal distribution of specific covariates in such a way that these components add up to the full composition effect. This is because most commonly used distributional features other than the mean, such as the variance, quantiles, or inequality measures, are nonlinear transformations of the distribution of the outcome variable, and are thus not additively separable in the marginal distributions of the covariates. Instead, they typically contain "interaction terms" that stem from the interplay of two or more covariates' marginal distributions, and are also influenced by between-group differences in the dependence patterns among the covariates. This is generally true even if the underlying data generating process does not contain any interaction terms itself. For example, when the data are generated by a simple linear model, between-group differentials in the outcomes' variances can be due to between-group differences in covariances among the explanatory variables. Since the covariance of two random variables is the product of their respective standard deviations and the correlation coefficient, it depends on features of both covariates' marginal distributions, and on the dependence structure between them. Such terms can not be apportioned unambiguously to specific covariates.

The main contribution of this paper is to derive a new decomposition of the composition effect that explicitly takes these issues into account. In particular, the decomposition contains three types of components: (i) the "direct contribution" of each covariate due to between-group differences in the respective marginal distributions, (ii) several "two way" and "higher order interaction effects" due to the interplay between two or more marginal distributions, and (iii) a "dependence effect" accounting for different dependence patterns among the covariates. We show that such a decomposition is well-defined in a general setting by using results from copula theory (e.g. Nelsen, 2006) and arguments related to those in Rothe (2012). Under suitable exogeneity conditions, these components can all be interpreted in terms of economically meaningful counterfactual experiments that change the joint distribution of the covariates to a hypothetical one that shares properties of both groups (Rothe, 2012).

To gain a better understanding of the empirical content of our decomposition, it is useful to consider a simple example. Suppose that we study the populations of workers in countries A and B, that our outcome variable is the hourly wage, and that there are two covariates: the worker's age and an indicator for union coverage. To be concrete, also suppose that population A tends to be older and more unionized, and that union coverage is more attractive for relatively old workers in population A than it is in population B. Then the composition effect is the part of the wage gap between the two populations that can be explained by these differences in observable characteristics. The first "direct contribution" is then the part of the composition effect that can be attributed to the fact that population A tends to be older. The second "direct contribution" accounts for population A having higher unionization rates. The (only) "interaction effect" measures the *additional* contribution of the fact that the population A is *both* more unionized and older. Finally, the "dependence effect" captures the fact that the relative age composition of union-covered and non-union-covered workers differs in the two populations.

An attractive feature of our decomposition is that it reduces to the Oaxaca-Blinder procedure for the special case of the mean of the outcome in a linear regression model. It can thus be understood as a natural extension of this method to nonlinear models and general distributional features. We also show in this paper that the elements of our decomposition are conceptually easy to estimate in a flexible parametric framework, requiring only standard statistical methods to estimate copula functions and conditional CDFs. The asymptotic variance of these estimates generally has a complicated form, but valid standard errors can be computed via a classical bootstrap approach.

As an additional contribution, we also apply our new methodology to study the evolution of the wage distribution among male workers in the United Stated between 1985 and 2005. There is now extensive evidence that during this period wage inequality in the United States has been rising substantially in the top end of the wage distribution, but has slightly decreased in the bottom end, leading to what is often called a polarization of the US labor market (Autor, Katz, and Kearney, 2006; Lemieux, 2008). Our methodology delivers some new and interesting insights into this issue. For example, our estimates suggest that our dependence effect, which has no analogue in other methods, accounts for about one fifth of the increase in wage inequality (as measured by the difference between the 90% and the 10% quantile) that occurred over the two decades, and about one fourth of the composition effect. This dependence effect captures for example changes in the positions of unionized and non-unionized workers in the age distribution.

It should be stressed that our focus in this paper is exclusively on a decomposition of the composition effect. We do not address the related issue of deriving a decomposition of the structure effect, i.e. dividing between-group differences in the structural functions that link the covariates and the outcome variable into components that can be attributed to individual covariates. Such a task already faces conceptional difficulties in simple linear models with discrete covariates, and does not seem possible for general nonlinear structural functions with interactions between the covariates.

The remainder of the paper is structured as follows. In the next section, we introduce a general setting for studying structure and composition effects, and illustrate the conceptional difficulties for decomposing the composition effect through a simple example. In Section 3, we describe our new decomposition based on copulas. Section 4 compares our approach to other approaches that have been proposed in the literature. Section 5 shows how to estimate the elements of our decomposition in practice, and Section 6 derives the asymptotic properties of these estimators. Section 7 contains an empirical application to wage data from the US. Finally, Section 8 concludes. Some proofs and more extensive calculations are given in the Appendix.

2. Structure and Composition Effects

2.1. General Setup. We consider a population with two non-overlapping subgroups indexed by $g \in \{0, 1\}$. For any individual in group g, we observe an outcome variable Y^g and a d-dimensional vector of observable characteristics X^g, with corresponding distribution functions F_Y^g and F_X^g, support $\mathcal{Y}^g$ and $\mathcal{X}^g$, and conditional CDF $F_{Y|X}^g$, respectively. Furthermore, for any CDF F we refer to objects of the form $\nu(F)$ as a distributional feature, where $\nu : \mathcal{F} \to \mathbb{R}$ is a functional from the space of all one-dimensional distribution functions to the real line. Examples of distributional features include the mean, with $\nu : F \mapsto \int y dF(y)$, and the τ-quantile, with $\nu : F \mapsto F^{-1}(\tau)$, but also higher-order centered or uncentered moments, quantile-related statistics like interquantile ranges or quantile ratios, and inequality measures such as the Gini coefficient. Our aim is to understand how the observed difference

$$\Delta_O^\nu = \nu(F_Y^1) - \nu(F_Y^0)$$

between the distributional features $\nu(F_Y^1)$ and $\nu(F_Y^0)$ is related to differences between the distributions F_X^1 and F_X^0. To this end, we define a counterfactual outcome distribution $F_Y^{g|j}$ that combines the conditional distribution in group g with the covariate distribution in group $j \neq g$:

$$F_Y^{g|j}(y) = \int F_{Y|X}^g(y, x) dF_X^j(x). \tag{2.1}$$

This integral is well-defined as long as $\mathcal{X}_j \subset \mathcal{X}_g$. We interpret $F_Y^{g|j}$ as the distribution of outcomes after a counterfactual experiment in which the distribution of observable characteristics in group g is changed from F_X^g to F_X^j, but the conditional distribution of the outcome given these characteristics remains constant. With the notation (2.1), one can decompose the observed between-group differential Δ_O^ν as

$$\Delta_O^\nu = \Delta_S^\nu + \Delta_X^\nu \tag{2.2}$$

where

$$\Delta_S^\nu = \nu(F_Y^1) - \nu(F_Y^{0|1}) \text{ and } \Delta_X^\nu = \nu(F_Y^{0|1}) - \nu(F_Y^0).$$

Here Δ_X^ν is a composition effect, solely due to differences in the distribution of the covariates between the two groups, and Δ_S^ν is a structure effect, solely due to differences in the conditional CDFs $F^1_{Y|X}$ and $F^0_{Y|X}$. This definition of structure and composition effects has been widely used in the literature (e.g. DiNardo et al., 1996; Machado and Mata, 2005; Rothe, 2010; Chernozhukov et al., 2013, among many others). We note that the order in which differences in conditional CDFs and the covariate distributions are considered when defining Δ^ν_X and Δ^ν_S has to be taken into account when interpreting the parameters in practice.

To give a concrete example that fits our setting, suppose that group 0 and 1 are the population of male workers in 1985 and 2005, respectively, Y^g is an individual's wage in constant 1985 dollars, X^g are observable characteristics that are relevant on the job market, and ν is some measure of inequality. In this case, the conditional CDF $F^g_{Y|X}$ summarizes the wage schedule given the worker's observable characteristics in period g, and $F^{g|j}_Y$ is the counterfactual distribution of wages that would have prevailed if the distribution of workers' characteristics in period g would have been the same as in period j. Moreover, Δ^ν_X and Δ^ν_S quantify to what extent changes in wage inequality over time, as measured by Δ^ν_O, can be attributed to the evolution of workers' characteristics and changes in the wage schedule, respectively.

Note that the structure and composition effects and can be given a causal interpretation under an (arguably strong) exogeneity conditions on the covariates. Specifically, suppose that the outcome variable is generated through the nonseparable model $Y^g = m^g(X^g, \eta^g)$ for $g \in \{0, 1\}$, where $\eta^g \in \mathbb{R}^{d_\eta}$ is an unobserved error term and m^g is the structural function. Now if η^g is independent of X^g then for any d-dimensional CDF G we can interpret the function $H_G(y) = \int F^g_{Y|X}(y, x)dG(x)$ as the CDF of the counterfactual random variable $m^g(Z, \eta^g)$, where $Z \sim G$ is a d-dimensional random vector that is independent of η^g.

2.2. Problems for Decomposing the Composition Effect. When the data contain information about several individual characteristics, it is natural to ask which role between-group differences in each one of them play in determining the composition effect. For example, one could be interested to which extent the composition part of the change in wage inequality from 1985 to 2005 can be attributed to the decline in unionization, the change in the distribution of workers' age, etc. Such questions can easily be addressed for the mean of the outcome variable in a simple linear model by the Oaxaca-Blinder procedure, which certainly contributes to the popularity of the method. This method does not apply, however, to other distributional features such as quantiles or variances, or more complex relationships between the outcome and the covariates.

It would thus be desirable to have a methodology that is able to apportion the composition effect into components attributable to between-group differences in the marginal distribution of each covariate in the general setting described above. However, it is clear that such a decomposition does generally not exist when the distributional feature of interest is a nonlinear transformation of the CDF of the outcome variable. This is the case for essentially all features commonly used in empirical applications, with the exception of the mean. These nonlinearities create issues even for simple data generating processes like linear models. To see this, consider the following simple example. Suppose that in both groups, i.e. for any $g \in \{0,1\}$, the data are generated as $Y^g = X_1^g + X_2^g + \eta^g$, where $X^g \sim N(\mu_g, \Sigma_g)$ is bivariate normal with

$$\mu_g = \begin{pmatrix} \mu_{g1} \\ \mu_{g2} \end{pmatrix} \text{ and } \Sigma_g = \begin{pmatrix} \sigma_{g1}^2 & \rho_g \sigma_{g1} \sigma_{g2} \\ \rho_g \sigma_{g1} \sigma_{g2} & \sigma_{g2}^2 \end{pmatrix},$$

and $\eta_g \sim N(0,1)$ is independent of X^g. Since the structural function is the same in both groups, we clearly have that $\Delta_S^\nu = 0$ for any functional ν, and thus any difference in the distribution of Y_1 and Y_0 reflects a composition effect. For instance, if $\nu(F_Y^g) = \text{Var}(Y^g)$ we have that $\Delta_X^\nu = \text{Var}(Y_1) - \text{Var}(Y_0)$ with $\text{Var}(Y^g) = 1 + \sigma_{g1}^2 + \sigma_{g2}^2 + 2\rho_g \sigma_{g1} \sigma_{g2}$. When $\nu(F_Y^g) = Q_Y^g(\tau)$ is the τ-quantile of the distribution of Y^g, we find that $\Delta_X^\nu = Q_Y^1(\tau) - Q_Y^0(\tau)$ with $Q_Y^g(\tau) = \mu_{g1} + \mu_{g2} + \Phi^{-1}(\tau)\sqrt{1 + \sigma_{g1}^2 + \sigma_{g2}^2 + 2\rho_g \sigma_{g1} \sigma_{g2}}$ and Φ the CDF of the standard normal distribution. In both cases, the object of interest is not additively separable in the parameters $(\mu_{g1}, \sigma_{g1}^2)$ and $(\mu_{g2}, \sigma_{g2}^2)$, which characterize the covariates' marginal distributions in this example. Moreover, in both cases the object of interest depends on the value of the correlation coefficient ρ_g, which determines the dependence structure among the covariates here.

3. Characterizing the Composition Effect using Copulas

3.1. Preliminaries. It turns out that the simple example in Section 2.2 illustrates a general property of the composition effect: it is usually not possible to express the composition effect as a sum of terms that each depend on the marginal distributions of a single covariate only. Instead, an explicit expression of the composition effect in terms of the respective marginal covariate distributions typically contains "interaction terms" resulting from the interplay of two or more marginal distributions, and also "dependence terms" resulting from between-group difference in the dependence pattern among the covariates.

To formally show this point, we use results from copula theory that allow us to disentangle the covariates' marginal distributions from the dependence structure among them. In particular, it follows from Sklar's Theorem (Sklar, 1959; Nelsen, 2006, Theorem 2.3.3) that the CDF of X^g can always be written as

$$F_X^g(x) = C^g(F_{X_1}^g(x_1), \ldots, F_{X_d}^g(x_d)) \text{ for } g \in \{0, 1\}, \tag{3.1}$$

where C^g is a copula function, i.e. a multivariate CDF with standard uniformly distributed marginals, and $F_{X_k}^g$ is the marginal distribution of the kth component of X^g. The copula describes the joint distribution of individuals' ranks in the various components of X^g, and can be interpreted as the object that determines the dependence structure.

To ensure that the definition of C^g is unique on its entire domain $[0, 1]^d$, we assume that the covariates X^g can be represented as $X^g = t(\tilde{X}^g) = (t_1(\tilde{X}_1^g), \ldots, t_d(\tilde{X}_d^g))$ for some continuously distributed random vector $\tilde{X}^g$ and a function $t(\cdot)$ that is weakly increasing in each of its arguments. For example, if the lth component of X^g is binary, we could have $X_l^g = \mathbb{I}\{\tilde{X}_l^g > c_l\} = t_l(\tilde{X}_l^g)$ for some constant c_l. If the lth component is continuously distributed, one can simply put t_l as the identity mapping. With such a structure, C^g can be uniquely defined on $[0, 1]^d$ as the copula function of the joint CDF of the latent variables $\tilde{X}^g$. While this construction uniquely defines C^g on $[0, 1]^d$, it does of course not ensure that C^g is point identified on $[0, 1]^d$ from the distribution of X^g alone when some of the covariates are discrete. We return to this issue in Section 3.4 below.

The representation (3.1) can be used to define counterfactual outcome distributions that combine the conditional distribution in group g with hypothetical covariate distributions that share properties of both F^1_X and F^0_X. Denoting any element of the d-dimensional product set $\{0,1\}^d$ by a boldface letter, we define the distribution of the outcome in a counterfactual setting where the structure is as in group g, the covariate distribution has the copula function of group j, and the marginal distribution of the lth covariate is equal to the that in group $\mathbf{k}_l$ by

$$F_Y^{g|j,\mathbf{k}}(y) = \int F^g_{Y|X}(y,x) dF_X^{j,\mathbf{k}}(x) \tag{3.2}$$

with

$$F_X^{j,\mathbf{k}}(x) \equiv C^j(F^{\mathbf{k}_1}_{X_1}(x_1), \ldots, F^{\mathbf{k}_d}_{X_d}(x_d)).$$

Following the discussion at the end of Section 2.1, under suitable exogeneity conditions this distribution can be interpreted as the one of the outcome variable in a counterfactual in which the distribution of the covariates is exogenously shifted such that its new CDF is equal to $F_X^{j,\mathbf{k}}$. We also write $\mathbf{1} = (1,1,\ldots,1)$ and $\mathbf{0} = (0,0,\ldots,0)$, denote by $\mathbf{e}^l$ the lth unit vector, i.e. the d-dimensional vector whose lth component is equal to one and whose remaining components are equal to zero, and put $|\mathbf{k}| = \sum_{l=1}^d \mathbf{k}_l$. Next, for any distributional feature ν we define the parameter

$$\beta^\nu(\mathbf{k}) = \nu(F_Y^{0|0,\mathbf{k}}) - \nu(F^0_Y),$$

which can be interpreted as the effect of a counterfactual experiment conducted in group 0 that changes the respective marginal distribution of those $|\mathbf{k}|$ covariates for which $\mathbf{k}_l = 1$ to their corresponding counterpart in group 1, while holding everything else (including the dependence structure among the covariates) constant. Note that the term $\beta^\nu(\mathbf{e}^l)$ is an example of a Fixed Partial Distributional Policy Effects (FPPE) as introduced in Rothe (2012). Finally, we define

$$\Delta^\nu_M(\mathbf{k}) = \beta^\nu(\mathbf{k}) + \sum_{1\leq|\mathbf{m}|\leq|\mathbf{k}|-1} (-1)^{|\mathbf{k}|-|\mathbf{m}|}\beta^\nu(\mathbf{m}),$$

with the empty sum equal to zero, so that e.g. $\Delta^\nu_M(\mathbf{e}^l) = \beta^\nu(\mathbf{e}^l)$. This notation will simplify the exposition below.

3.2. A Simple Example with Two Covariates. Before giving a general formula for our decomposition, it is instructive to first consider a simplified version where the vector of explanatory variables X^g has only two components, that is $d = 2$. In this case, we can rewrite the composition effect as follows:

$$\Delta_X^\nu = \Delta_M^\nu(\mathbf{e}^1) + \Delta_M^\nu(\mathbf{e}^2) + \Delta_M^\nu(\mathbf{1}) + \Delta_D^\nu \tag{3.3}$$

where

$$\Delta_D^\nu = \nu(F_Y^{0|1,\mathbf{1}}) - \nu(F_Y^{0|0,\mathbf{1}}) \quad \text{and} \quad \Delta_M^\nu(\mathbf{1}) = \beta^\nu(\mathbf{1}) - \beta^\nu(\mathbf{e}^1) - \beta^\nu(\mathbf{e}^2).$$

The representation (3.3) conveys a number of insights into the structure of the composition effect. First, the terms $\Delta_M^\nu(\mathbf{e}^1)$ and $\Delta_M^\nu(\mathbf{e}^2)$ can be interpreted as the *direct* contribution of between-group differences in the marginal distribution of the 1st and 2nd covariate to the composition effect, respectively. Second, the term $\Delta_M^\nu(\mathbf{1})$ can be interpreted as an interaction effect: while $\beta^\nu(\mathbf{1})$ measures the *joint* contribution of between-group differences in the marginal covariate distribution of between the two groups, subtracting $\beta^\nu(\mathbf{e}^1)$ and $\beta^\nu(\mathbf{e}^2)$ provides an adjustment for the *direct* contribution of the 1st and 2nd covariate, thus leading to the interpretation of $\Delta_M^\nu(\mathbf{1_2})$ as a "pure" interaction effect. Third, the term Δ_D^ν is a *dependence effect* that captures the contribution between-group differences in the covariates' copula functions. As a further illustration, in Appendix B we explicitly calculate the components of (3.3) for the simple setting described in Section 2.2. An example how these terms can be interpreted in an empirical context was given in the introduction of this paper.

3.3. A General Decomposition. We now describe a general decomposition of the composition effect for high-dimensional settings, which is the main contribution of this paper. As a first step, the composition effect Δ_X^ν can be decomposed into a *dependence effect* Δ_D^ν resulting from between-group differences in the copula functions, and a *total marginal distribution effect* Δ_M^ν resulting from differences in the marginal covariate distributions across the two groups:

$$\Delta_X^\nu = \Delta_D^\nu + \Delta_M^\nu \tag{3.4}$$

where

$$\Delta_D^\nu = \nu(F_Y^{0|1,\mathbf{1}}) - \nu(F_Y^{0|0,\mathbf{1}}) \text{ and } \Delta_M^\nu = \nu(F_Y^{0|0,\mathbf{1}}) - \nu(F_Y^{0|0,\mathbf{0}}).$$

Note that the order in which differences in the copula and the marginal CDFs are considered when defining Δ_D^ν and Δ_M^ν has to be taken into account when interpreting the parameters in practice.

In a second step, we further decompose the total marginal distribution effect Δ_M^ν into several *partial marginal distribution effects* $\Delta_M^\nu(\mathbf{k})$, which account for between-group differences in the marginal distributions of one or several covariates:

$$\Delta_M^\nu = \sum_{1\leq|\mathbf{k}|\leq d} \Delta_M^\nu(\mathbf{k}).$$

Note that the order of the covariates in the data set has no impact on the numerical value of each of the $\Delta_M^\nu(\mathbf{k})$, and that these terms have been carefully defined in the previous section in order to avoid counting certain contributions more than once.

Taken together the various terms we just introduced, our full decomposition of the composition effect is given by

$$\Delta_X^\nu = \sum_{1\leq|\mathbf{k}|\leq d} \Delta_M^\nu(\mathbf{k}) + \Delta_D^\nu. \tag{3.5}$$

In case $|\mathbf{k}| = 1$, i.e. when $\mathbf{k} = \mathbf{e}^l$ is the lth unit vector, the interpretation of $\Delta_M^\nu(\mathbf{e}^l)$ is, as described above, that of a direct contribution of between-group differences in the marginal distribution of the lth covariate to the composition effect. For any vector $\mathbf{k}$ with $|\mathbf{k}| > 1$, the terms $\Delta_M^\nu(\mathbf{k})$ capture the contributions to the composition effect of "$|\mathbf{k}|$-way interaction effects" between the marginal distributions for which respective component of $\mathbf{k}$ is equal to one.

It should be stressed that the decomposition (3.5) is not merely a statistical identity. As discussed in Section 2.1, under a suitable exogeneity condition on the covariates the terms Δ_D^ν and $\Delta_M^\nu(\mathbf{e}^l)$ for $l \in \{1, \ldots, d\}$ are all direct summary measures of the outcome of well-defined counterfactual experiments, namely that of a *ceteris paribus* change in the copula function or the lth marginal distribution, respectively. The terms of the form $\Delta_M^\nu(\mathbf{k})$ with $|\mathbf{k}| > 1$ arise from a comparison of the outcomes of several counterfactual experiments that all involve *ceteris paribus* changes of two or more covariates' marginal distributions. Thus all elements of the decomposition (3.5) can be interpreted in terms of empirically meaningful counterfactual experiments.

3.4. Identification. It is easy to see that the elements of the decomposition (3.5) are well-defined as long as the support of X_1 is contained in the support of X_0, i.e. $\mathcal{X}_1 \subset \mathcal{X}_0$. When some of the covariates are discrete, however, one generally requires further conditions on the data generating process in order to point-identify these quantities from the distribution of the observables. This is because the data identify the the copula function C^0 on the range of the respective marginal distribution functions of the components of X^0 only. There could thus be several copula functions $\tilde{C}$ that satisfy the relationship $F_X^0(x) = \tilde{C}(F_{X_1}^0(x_1), \ldots, F_{X_d}^0(x_d))$ for all $x \in \mathbb{R}^d$ if some of the components of X_0 are discrete (Nelsen, 2006). As a consequence, there might be no one-to-one mapping between the distribution of observables and the counterfactual distributional features of the form $\nu(F_Y^{0|0,\mathbf{s}})$ with $\mathbf{s} \in \{0,1\}^d$, which are required for computing our decomposition. Note that this is not an issue if all covariates are continuously distributed.

In the presence of discrete covariates, one way to achieve point identification of the elements of the decomposition is by imposing certain parametric restrictions on the functional form of the copula. These restrictions can be very mild though, and still allow for a wide rage of practically relevant dependence patterns. One might be willing to assume, for example, that C^0 belongs to the family of Gaussian copulas, or some other family that satisfies a weak regularity condition that we state below. By making such an assumption, one reduces the class of values that C^0 could potentially take such that it can be uniquely determined from the distribution of observables among the remaining candidates. The following proposition formalizes this argument.

Proposition 1. *Suppose that $\mathcal{X}_1 \subset \mathcal{X}_0$, and that either of the following conditions hold:*

(i) The distribution function F_X^0 is continuous.

(ii) The copula function $C^0 = C_{\theta^0}$ is contained in the parametric class $\{C_\theta, \theta \in \Theta \subset \mathbb{R}^k\}$, and it holds that $P(C_\theta(U_0) \neq C_{\theta^0}(U_0)) > 0$ for $U_0 = (F_{X_1}^0(X_{1,0}), \ldots, F_{X_d}^0(X_{d,0}))$ and all $\theta \in \Theta \setminus \{\theta_0\}$.

Then all terms in (3.5) *are point-identified.*

The statement of the proposition follows from elementary properties of copula functions (e.g. Nelsen, 2006) under condition (i), and from standard arguments for identification of parametric models (e.g. Newey and McFadden, 1994) under condition (ii). The former condition is obviously straightforward to verify, and the latter condition has testable implications. If the parametric family of copula functions is such that $C_\theta(u) \neq C_{\theta'}(u)$ for all $u \in (0,1)^d$ and all $\theta \neq \theta'$, then condition (ii) necessarily holds. This stronger sufficient condition can be shown to be fulfilled by most common parametric families of copula functions, such as e.g. the Gaussian, Clayton, or Frank family. In the absence of parametric restrictions, one could conduct a partial identification analysis similar to the one in Rothe (2012), and derive upper and lower bounds on the elements of the decomposition (3.5). Whether or not these bounds would be sufficiently narrow to be informative in an empirical setting depends on the details of the respective application.

4. Relationship to other Approaches

In this section, we compare our copula decomposition to several related approaches that have been proposed in the literature. We consider the classical Oaxaca-Blinder procedure, methods based on so-called sequential conditioning arguments, and a method based on RIF regression that was recently proposed by Firpo, Fortin, and Lemieux (2007; 2013).

4.1. Decompositions based on Linear Models. One important property of the decomposition (3.5) is that it encompasses the popular Oaxaca-Blinder procedure (Oaxaca, 1973; Blinder, 1973) as a special case. Suppose that the data are generated as

$$Y^g = \beta_0^g + \sum_{l=1}^{d} X_l^g \beta_l^g + \eta^g \text{ with } \mathbb{E}(\eta^g | X^g) = 0,$$

and the expectation is the distributional feature of interest, i.e. $\nu(F_Y^g) = \mathbb{E}(Y^g)$. Then it is straightforward to verify that the elements of our decomposition take the form

$$\Delta_M^{\mathbb{E}}(\mathbf{e}^l) = (\mathbb{E}(X_l^1) - \mathbb{E}(X_l^0))\beta_l^0,\ \Delta_M^{\mathbb{E}}(\mathbf{k}) = 0 \text{ for } \mathbf{k} \neq \mathbf{e}^l \text{ and } \Delta_D^{\mathbb{E}} = 0.$$

Our decomposition can thus be understood as a natural generalization of the Oaxaca-Blinder procedure to nonlinear DGPs and general features of the outcome distribution. Note that in this particular case the composition effect can be apportioned unambiguously into contributions attributable each covariate as the additive separability of covariates in the data generating process is preserved by the linearity of the functional $F \mapsto \int y dF(y)$ that maps a CDF into the corresponding expectation.

4.2. Decompositions based on Sequential Conditioning Arguments. Our decomposition does generally not encompass methods that are based on sequential conditioning arguments, such as those proposed by DiNardo et al. (1996), Machado and Mata (2005), Altonji et al. (2012) or Chernozhukov et al. (2013), for example. These approaches write the composition effect as a sum of terms that reflect changes in the *conditional* distribution of one covariate given the remaining ones. That is, they start with the representation

$$F_Y^g(y) = \int F_{Y|X}^g(y|x) dF_{X_1|X_{-1}}^g(x_1|x_{-1}) dF_{X_2|X_{-2}}^g(x_2|x_{-2}) \dots dF_{X_d}^g(x_d),$$

where a_{-k} denotes the subvector of a that excludes its first k components, and consider counterfactual distributions obtained by exchanging the various conditional covariate distributions with their counterparts in the other group. As argued in Rothe (2012), such terms cannot be interpreted as the impact of between-group differences in the marginal distribution of particular covariates. In contrast to our method, these decompositions are also *path dependent*, in the sense that they depend on the order of the components of X^g. This property can be undesirable in applications where there is no natural sequential order for the covariates.

4.3. Decompositions based on RIF Regressions. Our approach is related to the RIF decomposition that was recently proposed in Firpo et al. (2007; 2013). To explain the connection, assume for a moment that all covariates are continuously distributed. Let $RIF(Y^g, \nu)$ be the recentered influence function (Van der Vaart, 2000) of $\nu(F_Y^g)$, and let

$$h^{g,\nu}(x) \equiv \mathbb{E}(RIF(Y^g, \nu)|X^g = x)$$

be its conditional expectation function given X^g. Now define

$$\gamma^{g,\nu} = \mathbb{E}(\partial_x h^{g,\nu}(X^g)) \text{ and } \bar{\gamma}^{g,\nu} = \mathbb{E}(X^g X^{g\prime})^{-1} \mathbb{E}(X^g RIF(Y^g, \nu)),$$

which are the corresponding average derivative and the coefficients of a linear projection of $RIF(Y^g, \nu)$ onto X^g, respectively. The latter can be understood as an approximation to $\gamma^{g,\nu}$ that is easier to estimate in practice. The RIF decomposition of the composition effect is then given by

$$\Delta_X^\nu = \sum_{l=1}^{d} \Delta_{M,RIF}^\nu(l) + \Delta_{E,RIF}^\nu$$

where

$$\Delta_{M,RIF}^\nu(l) \equiv \bar{\gamma}_l^{0,\nu}(\mathbb{E}(X_l^1) - \mathbb{E}(X_l^0))$$

is interpreted as the contribution of the lth covariate and $\Delta_{E,RIF}^\nu$ is a residual term that is supposed to catch various types of misspecification.

Firpo, Fortin, and Lemieux (2009) show that the components of $\gamma^{g,\nu}$ measure the effect of infinitesimal location shifts in the marginal distribution of the corresponding component of X^g on $\nu(F_Y^g)$, holding everything else (including the copula) constant. The terms $\Delta_{M,RIF}^\nu(l)$ can thus be seen as "first order" approximations to our $\Delta_M^\nu(\mathbf{e}^l)$ along a particular path in the space of potential covariate distributions, namely the one described by a location shift in the lth margin. This approximation should be adequate if (i) the distribution of X_l^1 is close to a location-shifted version of the distribution of X_l^0, (ii) the location difference $\mathbb{E}(X_l^1) - \mathbb{E}(X_l^0)$ is small, and (iii) $\bar{\gamma}_l^{0,\nu}$ is close to $\gamma_l^{0,\nu}$. If one of these three conditions is violated then $\Delta_{M,RIF}^\nu(l)$ can differ from $\Delta_M^\nu(\mathbf{e}^l)$ by a potentially large amount.

In cases where the approximation is inaccurate, the elements of the RIF decomposition do generally not correspond to meaningful population quantities. For example, suppose that the distribution of X_l^1 and X_l^0 have the same mean but different variances, and are thus not location-shifted versions of each other. Then $\Delta_{M,RIF}^\nu(l)$ is clearly equal to zero by construction even though the lth covariate should in general contribute to the composition effect for any distributional feature ν that is affected by the spread of the covariate distribution. Note that checking whether the residual $\Delta_{E,RIF}^\nu$ is small does not ensure the accuracy of the RIF approximation since it is easy to construct examples where $\Delta_{E,RIF}^\nu = 0$ but $\Delta_{M,RIF}^\nu(l)$ differs from $\Delta_M^\nu(\mathbf{e}^l)$ for all l by an arbitrarily large amount.

5. Model Specification and Estimation

In this section, we explain how our detailed decomposition (3.4)–(3.5) can be estimated in practice. We focus on flexible parametric specifications, which can be estimated using standard statistical techniques. Recall that for any distributional feature ν the observed difference Δ_O^ν, the structure effect Δ_S^ν, the dependence effect Δ_D^ν, and the various partial marginal distribution effects $\Delta_M^\nu(\mathbf{k})$ can all be expressed in terms of objects of the form $\nu(F_Y^{g|j,\mathbf{k}})$, where

$$F_Y^{g|j,\mathbf{k}}(y) = \int F_{Y|X}^g(y,x)dC^j(F_{X_1}^{\mathbf{k}_1}(x_1),\ldots,F_{X_d}^{\mathbf{k}_d}(x_d))$$

as defined in equation (3.2). To obtain estimates of the elements of our decomposition, we can therefore use a plug-in approach and replace the terms $\nu(F_Y^{g|j,\mathbf{k}})$ at every occurrence with their sample counterparts $\nu(\widehat{F}_Y^{g|j,\mathbf{k}})$, where

$$\widehat{F}_Y^{g|j,\mathbf{k}}(y) = \int \widehat{F}_{Y|X}^g(y,x)d\widehat{C}^j(\widehat{F}_{X_1}^{\mathbf{k}_1}(x_1),\ldots,\widehat{F}_{X_d}^{\mathbf{k}_d}(x_d)), \tag{5.1}$$

with $\widehat{F}_{Y|X}^g$, $\widehat{C}^g$ and $\widehat{F}_{X_l}^g$ being suitable estimates of the conditional CDF $F_{Y|X}^g$, the copula function C^g, and the marginal CDFs $F_{X_l}^g$ of the lth component of X^g, for $g \in \{0,1\}$ and $m \in \{1,\ldots,d\}$, respectively. That is, our estimates are given by

$$\begin{aligned}
&\widehat{\Delta}_O^\nu = \nu(\widehat{F}_Y^{1|1,\mathbf{1}}) - \nu(\widehat{F}_Y^{0|0,\mathbf{0}}), \quad \widehat{\Delta}_S^\nu = \nu(\widehat{F}_Y^{1|1,\mathbf{1}}) - \nu(\widehat{F}_Y^{0|1,\mathbf{1}}),\\
&\widehat{\Delta}_X^\nu = \nu(\widehat{F}_Y^{0|1,\mathbf{1}}) - \nu(\widehat{F}_Y^{0|0,\mathbf{0}}), \quad \widehat{\Delta}_D^\nu = \nu(\widehat{F}_Y^{0|1,\mathbf{1}}) - \nu(\widehat{F}_Y^{0|0,\mathbf{1}}), \text{ and}\\
&\widehat{\Delta}_M^\nu(\mathbf{k}) = \nu(\widehat{F}_Y^{0|0,\mathbf{k}}) - \nu(\widehat{F}_Y^{0|0,\mathbf{0}}),
\end{aligned}$$

for any distributional feature ν. The integral on the right-hand side of equation (5.1) can be approximated numerically through any of the many standard methods available in commonly used software packages.

We assume that the data available to the econometrician consists of two i.i.d. samples $\{(Y_i^g, X_i^g)\}_{i=1}^{n_g}$ of size n_g from the distribution of (Y^g, X^g) for $g \in \{0,1\}$. In such a classical cross-sectional setting, estimates of the just-mentioned unknown functions can in principle be obtained by a variety of different methods. For the univariate distribution functions $F_{X_l}^g$, the most straightforward estimator is arguably the usual empirical CDF, which is given by

$$\widehat{F}_{X_l}^g(x_l) = \frac{1}{n_g}\sum_{i=1}^{n_g} \mathbb{I}\{X_{li}^g \le x_l\}.$$

For the higher-dimensional objects, i.e. the conditional CDFs and the copula functions, several nonparametric, semiparametric and fully parametric procedures have been proposed in the literature. Since most studies that make use of decomposition methods use data sets containing a large number of individual sociodemographic characteristics, the substantial sample size requirements of nonparametric methods in such high-dimensional settings limit their attractiveness. We therefore focus on flexible parametric specifications in this paper, which have been applied successfully in the empirical literature.

Our preferred method to estimate the conditional CDFs $F^g_{Y|X}$, which is also used in our empirical application below, is the distributional regression approach of Foresi and Peracchi (1995). The distributional regression model assumes that

$$F^g_{Y|X}(y,x) \equiv \Phi(x'\delta^g_o(y)),$$

where $\Phi(\cdot)$ is the standard normal CDF (or some other strictly increasing link function). The finite-dimensional parameter $\delta^g_o(y)$ can then be estimated by the maximum likelihood estimate $\widehat{\delta}^g(y)$ in a Probit model that relates the indicator variable $\mathbb{I}\{Y^g \leq y\}$ to the covariates X^g, i.e.

$$\widehat{\delta}^g(y) = \underset{\delta}{\operatorname{argmax}} \sum_{i=1}^{n_g} (\mathbb{I}\{Y^g_i \leq y\} \log(\Phi(X^{g\prime}_i\delta)) + (1 - \mathbb{I}\{Y^g_i \leq y\}) \log(1 - \Phi(X^{g\prime}_i\delta))).$$

The resulting estimate of the conditional CDF is then given by

$$\widehat{F}^g_{Y|X}(y,x) = \Phi(x'\widehat{\delta}^g(y)).$$

An alternative approach that is commonly found in the literature (e.g. Machado and Mata, 2005), would be to model the conditional quantile function by a linear quantile regression model (Koenker and Bassett, 1978; Koenker, 2005), and then invert the corresponding estimated quantile function. Chernozhukov et al. (2013) derive asymptotic properties of conditional CDF estimators based on distributional and quantile regression (and several other parametric specifications), establishing classical properties like $\sqrt{n_g}$-consistency and asymptotic normality under standard regularity conditions. Rothe and Wied (2013) consider specification testing in these types of models.

The copula functions C^g can also be modeled a variety of ways. Following the arguments preceding Proposition 1, we consider the case that C^g is contained in sufficiently flexible parametric class indexed by a k-dimensional parameter, i.e.

$$C^g \equiv C_{\theta^g_o} \in \{C_\theta, \theta \in \Theta \subset \mathbb{R}^k\}.$$

Different parametric copula models are able to generate different types of dependence patters, and thus the analyst should chose a specification that is considered flexible enough to encompass the relationship between the covariates in the respective group. The extensive reviews of the properties of various copula models in e.g. Nelsen (2006) or Trivedi and Zimmer (2007) are a useful guidance for this choice. When all covariates are continuously distributed, the dependence parameters of any common copula model can be estimated by Maximum Likelihood methods implemented in most common software packages, and estimates can be shown to be $\sqrt{n_g}$-consistent and asymptotically normal under standard regularity conditions (e.g. Genest, Ghoudi, and Rivest, 1995). When some of the covariates follow a discrete marginal distribution, Maximum Likelihood estimation is still conceptually possible, but might not be feasible as the computational burden of evaluating the likelihood function grows exponentially in the number of discrete covariates and their support points (Joe, 1997). In empirical settings with large samples and many discrete covariates, it is therefore more practical to use estimators for the copula parameters that are faster to compute, even though they might not be fully efficient. An example of an estimator that satisfies this criterion is the minimum distance estimator

$$\widehat{\theta}^g = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n_g} (\widehat{F}_X^g(X_{1i}^g, \ldots, X_{di}^g) - C_\theta(\widehat{F}_{X_1}^g(X_{1i}^g), \ldots \widehat{F}_{X_d}^g(X_{di}^g)))^2,$$

where $\widehat{F}_X^g$ denotes the joint empirical CDF of the covariate data from group g. Other estimators for copula functions with discrete margins have recently been proposed for example by Nikoloulopoulos and Karlis (2009), Smith and Khaled (2012) or Panagiotelis, Czado, and Joe (2012).

In our empirical application, which contains several covariates that are not expected to have the same pairwise dependence patters, we use the Gaussian copula model

$$C_\Sigma(u) = \Phi_\Sigma^d(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)),$$

with Φ_Σ^d the CDF of a d-variate standard normal distribution with correlation matrix Σ and Φ the standard normal CDF. This specification has a further computational advantage, namely that the (a, b) element of Σ only affects the pairwise dependence between X_a^g and X_b^g. We can therefore estimate the matrix Σ by $\widehat{\Sigma}$, whose (a, b) element is given by the "pairwise" minimum distance estimator

$$\widehat{\Sigma}_{(ab)} = \operatorname*{argmin}_{\rho} \sum_{i=1}^{n_g} (\widehat{F}^g_{X_a,X_b}(X^g_{ai}, X^g_{bi}) - \Phi^2_\rho(\Phi^{-1}(\widehat{F}^g_{X_a}(X^g_{ai})), \Phi^{-1}(\widehat{F}^g_{X_b}(X^g_{bi}))))^2$$

for all $1 \leq a < b \leq d$, where $\widehat{F}^g_{X_a,X_b}$ denotes the joint empirical CDF of X^g_a and X^g_b in group g, and Φ^2_ρ denotes a bivariate standard normal distribution with correlation ρ. This procedure avoids the high-dimensional numerical optimization of a full minimum distance estimator, and is thus much faster and stable from a computational point of view than a standard minimum distance estimator.

6. Asymptotic Theory

In this section, we derive some asymptotic properties of the estimators of the components of our decomposition introduced above. These will allow us to conduct inference. We start by introducing the following assumptions.

Assumption 1 (Sampling). *For $g \in \{0, 1\}$ the data $\{(Y^g_i, X^g_i)\}^{n_g}_{i=1}$ are an independent and identically distributed sample from the distribution of (Y_g, X_g).*

Requiring simple random sampling within the two populations is standard for the types of microeconometric applications that we have in mind. However, our approach could in principle be extended to settings with certain types of non-independent observations, such as time series or clustered data.

Assumption 2 (CDF Estimator). *The estimator $\widehat{F}^g_{Y|X}$ is such that as $n_g \to \infty$*

$$\sqrt{n_g}(\widehat{F}^g_{Y|X}(y, x) - F^g_{Y|X}(y, x)) = \frac{1}{\sqrt{n_g}} \sum_{i=1}^{n_g} \psi^g_i(y, x) + o_P(1),$$

uniformly over $(y, x) \in \mathcal{YX}$, with $\psi^g_i(y, x)$ a function of (Y^g_i, X^g_i) that is such that $\mathbb{E}(\psi^g_i(y, x)) \equiv 0$ and $\sup_{(y,x)\in\mathcal{YX}} \mathbb{E}(\psi^g_i(y, x)^2) < \infty$, for $g \in \{0, 1\}$.

This assumption requires the estimator of the conditional CDF of Y^g given X^g to be asymptotically linear and to converge at the standard parametric rate $\sqrt{n_g}$. It is a "high level" condition that can be shown to be satisfied more many standard parametric estimators under mild regularity conditions on the primitives of the data generating process.

Chernozhukov et al. (2013) derive such a result for the above-mentioned estimator based on the distributional regression model $F^g_{Y|X}(y,x) = \Phi(x'\delta^g_o(y))$, which is our preferred specification, showing that it holds with

$$\psi^g_i(y,x) = \phi(x\delta^g_o)x\frac{\mathbb{I}\{Y^g_i \leq y\} - \Phi(X^{g\prime}_i\delta^g_o)}{\Phi(X^{g\prime}_i\delta^g_o)(1-\Phi(X^{g\prime}_i\delta^g_o))}\phi(X^{g\prime}_i\delta^g_o)X^g_i$$

with ϕ the derivative of Φ, under some mild regularity conditions. Chernozhukov et al. (2013) also consider several other approaches, including linear quantile regression.

Assumption 3 (Copula Estimator)**.** *The copula function $C^g \equiv C_{\theta^g_o}$ is contained in the parametric class $\{C_\theta, \theta \in \Theta \subset \mathbb{R}^k\}$, which satisfies condition (ii) of Proposition 1. Moreover, the estimator $\widehat{\theta}^g$ is such that as $n_g \to \infty$*

$$\sqrt{n_g}(\widehat{\theta}^g - \theta^g_o) = \frac{1}{\sqrt{n_g}}\sum_{i=1}^{n_g}\xi^g_i + o_P(1),$$

with ξ^g_i a function of (Y^g_i, X^g_i) that is such that $\mathbb{E}(\xi^g_i) = 0$ and $\mathbb{E}(\xi^g_i\xi^{g\prime}_i) < \infty$, for $g \in \{0,1\}$

Assumption 3 is again an asymptotic linearity condition, this time on the estimate of the Copula parameter θ^g_o. This type of condition is standard for parametric estimation. Sufficient conditions for this assumption to hold for extremum estimators like Maximum Likelihood or the minimum distance approach described above can e.g. be found in Newey and McFadden (1994).

Assumption 4 (Smoothness)**.** *(i) The elements of the parametric class $\{C_\theta(x), \theta \in \Theta \subset \mathbb{R}^k\}$ of copula functions are continuously differentiable with respect to x and θ. The corresponding derivatives $c^g(x) = \partial_x C_{\theta^g_o}(x)$ and $C^{g,\theta}(x) = \partial_\theta C_{\theta^g_o}(x)$ are uniformly bounded. (ii) The CDF $F^g_{Y|X}(y,x)$ is continuously differentiable with respect to x, and the derivative is uniformly bounded. (iii) The functional ν is Hadamard differentiable on $\mathcal{F}$ with derivative ν'.*

This assumption imposes some weak smoothness conditions on the Copula functions, the conditional CDFs, and the distributional feature of interest. These are necessary for Delta Method-type arguments to apply. We remark that most commonly considered distributional features of interest are Hadamard differentiable under standard conditions, including moments, quantiles and many types of inequality measures. See Rothe (2010) for further examples and explicit formulas for the derivatives.

The following proposition considers the asymptotic properties of objects of the form $\nu(\widehat{F}_Y^{g|j,\mathbf{k}})$ in an asymptotic setting where the relative size of the two samples remains constant as both their absolute sizes tend to infinite, i.e. $n_g/(n_1+n_0) = \lambda_g$ for some fixed positive constants λ_1, λ_0 as $n_1 \to \infty$ and $n_0 \to \infty$. As a final piece of notation, let $(\xi^1, \psi^1, X^1, \xi^0, \psi^0, X^0)$ be a random vector such that $\{(\xi_i^g, \psi_i^g, X_i^g)\}_{i=1}^{n_g}$ is an i.i.d. sample from the distribution of (ξ^g, ψ^g, X^g) for $g \in \{0,1\}$. Note that such a vector always exists under our assumptions, and that it is such that (ξ^0, ψ^0, X^0) and (ξ^1, ψ^1, X^1) are stochastically independent. We then obtain the following result (see the Appendix for a formal proof).

Proposition 2. *Under Assumption 1–4,*

$$\sqrt{n}(\nu(\widehat{F}_Y^{g|j,\mathbf{k}}) - \nu(F_Y^{g|j,\mathbf{k}})) \xrightarrow{d} \mathbb{G}^{g|j,\mathbf{k}} \cdot \nu'(\zeta^{g|j,\mathbf{k}})$$

jointly over $(g,j,\mathbf{k}) \in \{0,1\} \times \{0,1\} \times \{0,1\}^d$ *as* $n \equiv n_1 + n_0 \to \infty$*, where*

$$\{\mathbb{G}^{g|j,\mathbf{k}} : (g,j,\mathbf{k}) \in \{0,1\} \times \{0,1\} \times \{0,1\}^d\}$$

is a collection of standard Gaussian random variables that are mutually independent and independent of the data, and

$$\begin{aligned}\zeta^{g|j,\mathbf{k}}(y) = &\sum_{s=1}^{d} \lambda_{\mathbf{k}_s}^{-1/2} \int F_{Y|X}^g(y|x)(c^j(F_X^{\mathbf{k}}(X_s^{\mathbf{k}_s}, x_{-s})) - c^j(F_X^{\mathbf{k}}(x)))dF_{X_1}^{\mathbf{k}_1}(x_1)\ldots dF_{X_d}^{\mathbf{k}_d}(x_d) \\ &+ \lambda_g^{-1/2} \int \psi^g(y,x)dF_X^{j,\mathbf{k}}(x) + \lambda_j^{-1/2}\xi^j \int F_{Y|X}^g(y|x)dC^{j,\theta}(F_X^{\mathbf{k}}(x)),\end{aligned}$$

where $C^{j,\theta}(F_X^{\mathbf{k}}(x)) = C^{j,\theta}(F_X^{\mathbf{k}_1}(x_1), \ldots, F_X^{\mathbf{k}_d}(x_d))$*,* $F_X^{j,\mathbf{k}}(x) = C^j(F_{X_1}^{\mathbf{k}_1}(x_1), \ldots, F_{X_d}^{\mathbf{k}_d}(x_d))$*, and* $c^j(F_X^{\mathbf{k}}(X_s^{\mathbf{k}_s}, x_{-s}) = c^j(F_{X_1}^{\mathbf{k}_1}(x_1), \ldots, F_{X_s}^{\mathbf{k}_s}(X_{\mathbf{k}_s,s}), \ldots, F_{X_d}^{\mathbf{k}_d}(x_d))$*.*

The proposition shows the distribution of any random vector with generic element $\sqrt{n}(\nu(\widehat{F}_Y^{g|j,\mathbf{k}}) - \nu(F_Y^{g|j,\mathbf{k}}))$ converges to a mean zero multivariate normal distribution, with the asymptotic covariance between two generic elements indexed by $(g,j,\mathbf{k})$ and $(g^*, j^*, \mathbf{k}^*)$ given by $\mathbb{E}(\nu'(\zeta^{g|j,\mathbf{k}})\nu'(\zeta^{g^*|j^*,\mathbf{k}^*}))$. Since all elements of our detailed decomposition take the form of linear combinations of these types of objects, the proposition implies that they are (jointly) asymptotically normal as well. For example, for the term $\widehat{\Delta}_M^{\nu}(\mathbf{e}^l) = \nu(\widehat{F}_Y^{0|0,\mathbf{e}^l}) - \nu(\widehat{F}_Y^{0|0,\mathbf{0}})$ the proposition implies that

$$\sqrt{n}(\widehat{\Delta}_M^{\nu}(\mathbf{k}) - \Delta_M^{\nu}(\mathbf{k})) \xrightarrow{d} N(0, \mathbb{E}((\nu'(\zeta^{0|0,\mathbf{e}^l}) - \nu'(\zeta^{0|0,\mathbf{0}}))^2)).$$

Similar results hold for the other elements of the decomposition. Since the asymptotic variance of these objects takes a fairly complicated form, in practice the most convenient way to estimate them seems to be a standard nonparametric bootstrap procedure in which the estimates are re-computed a large number of times on bootstrap samples $\{(\widetilde{Y}_i^g, \widetilde{X}_i^g)\}_{i=1}^{n_g}$ drawn with replacement from the original data $\{(Y_i^g, X_i^g)\}_{i=1}^{n_g}$, for $g \in \{0, 1\}$. The bootstrap variance estimator then coincides with the empirical variance of the bootstrap estimates. Such an approach is valid under conditions of the proposition as long as the obvious analogues of Assumption 2 and Assumption 3 also hold for the bootstrap estimates of the CDF and the copula parameters. Conditions for the former can e.g. be found in Chernozhukov et al. (2013), and are standard for the latter.

7. An Empirical Application

In this section, we provide a small-scale empirical study that illustrates the application of our decomposition of the composition effect in practice. Using data from the Current Population Survey (CPS), we decompose differences in various features of the 1985 and 2005 distribution of wages among male workers in the United States. There is now extensive evidence that during this period wage inequality in the United States has been rising substantially in the top end of the wage distribution, but has slightly decreased in the bottom end, leading to what is often called a polarization of the US labor market (Autor et al., 2006; Lemieux, 2008). The main new insight delivered by our methodology is that the dependence effect can explain a substantial proportion of the increase in overall wage inequality. This is an interesting finding, since the dependence effect has no immediate analogue in other decomposition methods that are typically used for this type of data.

We use a data set from Fortin et al. (2011), which was extracted from the 1983–1985 and 2003–2005 Outgoing Rotation Group (ORG) supplements of the CPS. See Lemieux (2006) for further details on its construction. Our data contain information on 232,784 and 170,693 males, respectively, that were employed in the relevant periods. Workers in the 1983–1985 and 2003–2005 sample play the role of our group 0 and 1. The outcome variable of interest is the log hourly wage, measured in constant 1985 dollars and multiplied by 100 (to improve the readability of the tables below). The covariates are years of education, years of potential labor market experience, and dummies for union coverage, race, marital status, and part-time status. All observations are weighted by the product of the number of hours worked and their respective CPS sample weights.

[Table 1 about here.]

Some descriptive statistics are given in Table 1. Results on wages confirm the general picture that was found before in the literature. Over the sample period, the 90% quantile of the wage distribution has been rising substantially, while the 10% quantile and the median exhibit only a moderate increase or have remained approximately constant, respectively. There is thus a large increase in wage inequality as measured by the difference between the 90% and the 10% quantile, but this is the effect of an increase in inequality in the right tail of the distribution. Regarding the covariates, the most striking feature is certainly the decline in union coverage from 27% in 1983–1985 to only 15% in 2003–2005. The average (potential) labor market experience and the average years of education both increased substantially by more than two years and six months, respectively. Changes in other explanatory variables are less pronounced.

To estimate the various elements of our decomposition of the composition effect, we proceed as described in Section 5. We model the copula functions C^g by a Gaussian copula, and the conditional CDFs $F^g_{Y|X}$ by a distributional regression model with a Gaussian link function. Compared to an approach based on quantile regression, distributional regression has the advantage that it is not affected by heaping in the distribution of wages, and seems to be better suited to capture the somewhat irregular behavior of the conditional wage distribution around the level of the minimum wage (Chernozhukov et al., 2013; Rothe and Wied, 2013). In addition to the covariates mentioned above, we use quadratic terms in education and experience and a full set of interaction terms for estimating the conditional CDFs. Standard errors are calculated via the nonparametric bootstrap, using $B = 200$ replications. Due to the large sample sizes, sampling variation in our estimates is mostly negligible.

[Table 2 about here.]

[Table 3 about here.]

Table 2 and Table 3 present the results of our decomposition for various measures of location and spread, respectively. Row by row, we report estimates of the total change Δ_O^ν, the usual structure and composition effect Δ_S^ν and Δ_X^ν, our dependence and marginal distribution effects Δ_D^ν and Δ_M^ν, the direct contributions $\Delta_M^\nu(\mathbf{e}^l)$ for each of the six covariates, and "two-way" interaction terms $\Delta_M^\nu(\mathbf{k})$ with $|\mathbf{k}| = 2$. For brevity, we do not report estimates higher-order interactions. Estimates of the total change Δ_O^ν, which impose the parametric restrictions on copulas and conditional distribution functions as explained above, are very close to the respective descriptive statistics that can be calculated directly from Table 1. We take this as an assuring indication that our parametric model provides a reasonable fit.

We first consider estimates of the structure and composition effect, which are both parameters that could also be estimated via other methods (e.g. DiNardo et al., 1996; Machado and Mata, 2005; Rothe, 2010; Chernozhukov et al., 2013, among many others). Table 2 shows that changes in labor force composition alone can explain a substantial part of the total change in the mean wage, but they do not offer an explanation for the differential change at various quantiles. For example, the estimates suggest that changes in labor force composition alone would have lead to a large upward shift of the median wage, while the total observed change between 1985 and 2005 was in fact close to zero. We can also see from Table 2 that the composition effect is positive for each of the three quantiles under consideration, with the magnitude of the effect gradually increasing with the quantile level. Correspondingly, the composition effect amount to about two thirds of the increase in the 90%-10% quantile differences, but has the opposite sign of the total change in the 50%-10% quantile differences.

We now consider the estimates of the further decomposition proposed in this paper, starting with the dependence effect. To interpret this component, it is important to understand the changes in the dependence structure of the covariates that took place over the study period. To that end, Table 4 reports estimates of the copula parameters from the 1985 and 2005 samples, respectively (recall that we are working with a Gaussian copula that is parametrized through a correlation matrix).

[Table 4 about here.]

The point estimates show several interesting developments. For example, negative associations between union coverage and education as well as being nonwhite and education essentially vanished over the study period. Working part-time became more negatively associated with education. Being married became more strongly associated with education, and somewhat less strongly with experience. The dependence effect measures the joint contribution of these changes. From Table 3, we can see that it plays a substantial role for the various measures of spread, explaining about 25% of the respective composition effects. In case of the 10% quantile, the dependence effect is of roughly the same absolute magnitude as the composition effect but has the opposite sign. For the other measures of location in Table 2, the dependence effect is relatively small.

Next, we consider the "direct" marginal distribution effects and "two-way" interaction effects of the six explanatory variables. From Table 2, we find that changes in the distribution of education and experience both have a similar, strongly positive "direct" impact on all quantiles, with slightly larger magnitudes for the median relative to the two extreme quantiles. That is, our estimates suggested that a counterfactual change in the marginal distribution of either education or experience to its respective 2005 value would have shifted the 1985 wage distribution to the right, while otherwise approximately preserving its overall shape. Correspondingly, Table 3 shows that changes in education or experience alone contribute only moderately to the total increase in inequality observed over that period. Moreover, we see from Table 2 that these two variables have a sizable negative interaction effect on the median, and a positive interaction effect of similar absolute magnitude on the 10% quantile. As a consequence, we also observe corresponding nonzero interaction terms for the various quantile differences in Table 3. To better understand the interpretation of the interaction effects, consider the case of the median. Our estimates suggest that a counterfactual change in the marginal distribution of *both* education or experience to its respective 2005 value (while holding everything else constant) would have increased the 1985 median by the sum of the two direct contributions and the corresponding interaction term, i.e. by $5.631 + 5.629 - 0.966 = 10.294$.

The results also suggest that the decline in unionization had an important impact on the distribution of wages. Changes in union coverage rates are estimated to have a strong negative effect on the median wage, a somewhat less negative one on the 10% quantile, and a small positive effect on the 90% quantile. This covariate is thus the single most important one for explaining changes in overall and top-end inequality, as it accounts for about 25% of the observed change in each the 90%-10% and 90%-50% quantile differences alone. There are also some minor interactions between union coverage and both education and experience.

Finally, our estimates suggest that changes in the proportion of part-time, non-white, and married workers are of comparatively minor importance relative to the aforementioned covariates when it comes to explaining the composition effect. This comes as no surprise as Table 1 shows that the marginal distribution of these variables did not change much between 1985 and 2005.

8. Conclusions

Studying the role of specific covariates in determining between-group differences in economic outcomes involves a number of subtle yet important issues. Using results from copula theory, we show that the composition effect naturally has an interesting structure that can be exploited in empirical applications. In particular, we show that the composition effect can be written as the sum of three types of components: (i) the "direct contribution" of each covariate due to between-group differences in the respective marginal distributions, (ii) several "two way" and "higher order interaction effects" due to the interplay between two or more marginal distributions, and (iii) a "dependence effect" accounting for different dependence patterns among the covariates. We show how these components can be estimated using flexible parametric specifications, and illustrate the procedure through an application to US wage data. The empirical application suggests that our method is able to uncover new and interesting features of the data, that have thus far not been detected by other approaches.

A. Proof of Asymptotic Properties

In this section, we formally prove the asymptotic normality result given in Section 6. For notational simplicity, suppose that $n_1 = n_0 \equiv n_*$, which means that we can work with

$\lambda_1 = \lambda_0 = 1/2$. As a first step, we write the object of interest $\widehat{F}_Y^{g|j,\mathbf{k}}(y)$ as

$$\begin{aligned}
\widehat{F}_Y^{g|j,\mathbf{k}}(y) &= \int \widehat{F}_{Y|X}^{g}(y,x) d\widehat{C}^j(\widehat{F}_{X_1}^{\mathbf{k}_1}(x_1), \ldots, \widehat{F}_{X_d}^{\mathbf{k}_d}(x_d)) \\
&= \int (\widehat{F}_{Y|X}^{g}(y,x) - F_{Y|X}^{g}(y,x)) d\widehat{C}^j(\widehat{F}_{X_1}^{\mathbf{k}_1}(x_1), \ldots, \widehat{F}_{X_d}^{\mathbf{k}_d}(x_d)) \\
&\quad + \int F_{Y|X}^{g}(y,x) d\widehat{C}^j(\widehat{F}_{X_1}^{\mathbf{k}_1}(x_1), \ldots, \widehat{F}_{X_d}^{\mathbf{k}_d}(x_d)) \\
&\equiv T_1 + T_2.
\end{aligned}$$

Using Assumption 2–3 and the Glivenko-Cantelli Theorem, it follows that

$$T_1 = \frac{1}{n_*} \sum_{i=1}^{n_*} \int \psi_i^g(y,x) dF_X^{j,\mathbf{k}}(x) + o_P(n_*^{-1/2})$$

uniformly over y. Next, we write T_2 as

$$\begin{aligned}
T_2 &= \int F_{Y|X}^{g}(y,x) d(\widehat{C}^j(\widehat{F}_{X_1}^{\mathbf{k}_1}(x_1), \ldots, \widehat{F}_{X_d}^{\mathbf{k}_d}(x_d)) - C^j(\widehat{F}_{X_1}^{\mathbf{k}_1}(x_1), \ldots, \widehat{F}_{X_d}^{\mathbf{k}_d}(x_d))) \\
&\quad + \int F_{Y|X}^{g}(y,x) dC^j(\widehat{F}_{X_1}^{\mathbf{k}_1}(x_1), \ldots, \widehat{F}_{X_d}^{\mathbf{k}_d}(x_d)) \\
&\equiv T_3 + T_4.
\end{aligned}$$

Using Assumption 3–4 and the Glivenko-Cantelli Theorem, it follows that

$$\begin{aligned}
T_3 &= -(\widehat{\theta}^j - \theta_o^j) \int F_{Y|X}^{g}(y,x) dC^{j,\theta}(F_X^{\mathbf{k}}(x)) + o_P(n^{-1/2}) \\
&= -\frac{1}{n_*} \sum_{i=1}^{n_*} \xi_i^j \int F_{Y|X}^{g}(y,x) dC^{j,\theta}(F_X^{\mathbf{k}}(x)) + o_P(n^{-1/2}).
\end{aligned}$$

uniformly over y. Finally, the term T_4 can be written as

$$\begin{aligned}
T_4 &= \int F_{Y|X}^{g}(y,x) c^j(\widehat{F}_X^{\mathbf{k}}(x)) d\widehat{F}_{X_1}^{\mathbf{k}_1}(x_1) \ldots d\widehat{F}_{X_d}^{\mathbf{k}_d}(x_d)) \\
&= \int F_{Y|X}^{g}(y,x) c^j(F_X^{\mathbf{k}}(x)) d(\widehat{F}_{X_1}^{\mathbf{k}_1}(x_1) - F_{X_1}^{\mathbf{k}_1}(x_1)) dF_{X_2}^{\mathbf{k}_2}(x_2), \ldots, dF_{X_d}^{\mathbf{k}_d}(x_d)) + \ldots \\
&\quad + \int F_{Y|X}^{g}(y,x) c^j(F_X^{\mathbf{k}}(x)) dF_{X_1}^{\mathbf{k}_1}(x_1) \ldots dF_{X_{d-1}}^{\mathbf{k}_{d-1}}(x_{d-1}) d(\widehat{F}_{X_d}^{\mathbf{k}_d}(x_d) - F_{X_d}^{\mathbf{k}_1}(x_d)) \\
&\quad + F_Y^{g|j\mathbf{k}}(y) + o_P(n^{-1/2}) \\
&= \frac{1}{n^*} \sum_{s=1}^{d} \sum_{i=1}^{n^*} \int F_{Y|X}^{g}(y|x)(c^j(F_X^{\mathbf{k}}(X_{si}^{\mathbf{k}_s}, x_{-s})) - c^j(F_X^{\mathbf{k}}(x))) dF_{X_1}^{\mathbf{k}_1}(x_1) \ldots dF_{X_d}^{\mathbf{k}_d}(x_d) \\
&\quad + F_Y^{g|j\mathbf{k}}(y) + o_P(n^{-1/2})
\end{aligned}$$

uniformly over y. Taken together, we have thus shown that

$$\widehat{F}_Y^{g|j,\mathbf{k}}(y) - F_Y^{g|j\mathbf{k}}(y) = \frac{1}{n^*}\sum_{i=1}^{n^*} \zeta_i^{g|j,\mathbf{k}}(y) + o_P(n^{-1/2})$$

uniformly over y, where

$$\begin{aligned}\zeta_i^{g|j,\mathbf{k}}(y) = &\sum_{s=1}^{d}\int F_{Y|X}^g(y|x)(c^j(F_X^{\mathbf{k}}(X_{si}^{\mathbf{k}_s}, x_{-s})) - c^j(F_X^{\mathbf{k}}(x)))dF_{X_1}^{\mathbf{k}_1}(x_1)\dots dF_{X_d}^{\mathbf{k}_d}(x_d)\\ &+\int \psi_i^g(y,x)dF_X^{j,\mathbf{k}}(x) - \xi_i^j\int F_{Y|X}^g(y,x)dC^{j,\theta}(F_X^{\mathbf{k}}(x)).\end{aligned}$$

The statement of the proposition then follows from the Functional Central Limit Theorem and the Functional Delta Method. □

B. A Simple Example

In this section, we explicitly calculate the components of our decomposition of the composition effect for the simple setting described in Section problems. Recall that in group $g \in \{0,1\}$ the data are generated as $Y^g = X_1^g + X_2^g + \eta^g$, the covariates $X^g \sim N(\mu_g, \Sigma_g)$ are bivariate normal with

$$\mu_g = \begin{pmatrix}\mu_{g1}\\ \mu_{g2}\end{pmatrix} \text{ and } \Sigma_g = \begin{pmatrix}\sigma_{g1}^2 & \rho_g\sigma_{g1}\sigma_{g2}\\ \rho_g\sigma_{g1}\sigma_{g2} & \sigma_{g2}^2\end{pmatrix},$$

and $\eta^g \sim N(0,1)$. In this setting, the the copula C^g of the distribution of X^g is the bivariate Gaussian copula with one-dimensional parameter ρ_g. For the case that $\nu(F_Y^g) = \text{Var}(Y^g)$, we find that the dependence structure and total marginal distribution effects are

$$\begin{aligned}\Delta_D^{\text{Var}} &= 2(\rho_1 - \rho_0)\sigma_{11}\sigma_{12} \text{ and}\\ \Delta_M^{\text{Var}} &= (\sigma_{11}^2 - \sigma_{01}^2) + (\sigma_{12}^2 - \sigma_{02}^2) + 2\rho_0(\sigma_{11}\sigma_{12} - \sigma_{01}\sigma_{02}),\end{aligned}$$

respectively, and that the partial marginal distribution effects are

$$\begin{aligned}\Delta_M^{\text{Var}}(\mathbf{e}^1) &= (\sigma_{11}^2 - \sigma_{01}^2) + 2\rho_0\sigma_{02}(\sigma_{11} - \sigma_{01}),\\ \Delta_M^{\text{Var}}(\mathbf{e}^2) &= (\sigma_{12}^2 - \sigma_{02}^2) + 2\rho_0\sigma_{01}(\sigma_{12} - \sigma_{02}), \text{ and}\\ \Delta_M^{\text{Var}}(\mathbf{1}) &= -2\rho_0(\sigma_{01}\sigma_{12} + \sigma_{02}\sigma_{11}).\end{aligned}$$

For $\nu(F_Y^g) = Q_Y^g(\tau)$ the distributional feature of interest, we find that

$$\Delta_D^{Q(\tau)} = \Phi^{-1}(\tau)\left(\sqrt{1+\sigma_{11}^2+\sigma_{12}^2+2\rho_1\sigma_{11}\sigma_{12}} - \sqrt{1+\sigma_{11}^2+\sigma_{12}^2+2\rho_0\sigma_{11}\sigma_{12}}\right) \text{ and}$$

$$\Delta_M^{Q(\tau)} = \Phi^{-1}(\tau)\left(\sqrt{1+\sigma_{11}^2+\sigma_{12}^2+2\rho_0\sigma_{11}\sigma_{12}} - \sqrt{1+\sigma_{01}^2+\sigma_{02}^2+2\rho_0\sigma_{01}\sigma_{02}}\right)$$
$$+ (\mu_{11}+\mu_{12}) - (\mu_{01}+\mu_{02}),$$

and that the partial marginal distribution effects are

$$\Delta_M^{Q(\tau)}(\mathbf{e}^1) = \Phi^{-1}(\tau)\left(\sqrt{1+\sigma_{11}^2+\sigma_{02}^2+2\rho_0\sigma_{11}\sigma_{02}} - \sqrt{1+\sigma_{01}^2+\sigma_{02}^2+2\rho_0\sigma_{01}\sigma_{02}}\right)$$
$$+ (\mu_{11}-\mu_{01}),$$
$$\Delta_M^{Q(\tau)}(\mathbf{e}^2) = \Phi^{-1}(\tau)\left(\sqrt{1+\sigma_{01}^2+\sigma_{12}^2+2\rho_0\sigma_{01}\sigma_{12}} - \sqrt{1+\sigma_{01}^2+\sigma_{02}^2+2\rho_0\sigma_{01}\sigma_{02}}\right)$$
$$+ (\mu_{12}-\mu_{02}), \text{ and}$$
$$\Delta_M^{Q(\tau)}(\mathbf{1}) = \Phi^{-1}(\tau)\left(\sqrt{1+\sigma_{11}^2+\sigma_{12}^2+2\rho_0\sigma_{11}\sigma_{12}} + \sqrt{1+\sigma_{01}^2+\sigma_{02}^2+2\rho_0\sigma_{01}\sigma_{02}}\right)$$
$$- \Phi^{-1}(\tau)\left(\sqrt{1+\sigma_{11}^2+\sigma_{02}^2+2\rho_0\sigma_{11}\sigma_{02}} + \sqrt{1+\sigma_{01}^2+\sigma_{12}^2+2\rho_0\sigma_{01}\sigma_{12}}\right)$$

Other distributional features could be considered in a similar fashion.

References

Altonji, J., P. Bharadwaj, and F. Lange (2012): "Changes in the characteristics of american youth: Implications for adult outcomes," *Journal of Labor Economics*, 40, 783–828.

Autor, D., L. Katz, and M. Kearney (2006): "The polarization of the US labor market," *American Economic Review*, 96, 189–194.

Barsky, R., J. Bound, K. Charles, and J. Lupton (2002): "Accounting for the black-white wealth gap," *Journal of the American Statistical Association*, 97, 663–673.

Blinder, A. (1973): "Wage discrimination: reduced form and structural estimates," *Journal of Human Resources*, 8, 436–455.

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2013): "Inference on Counterfactual Distributions," *Econometrica*, 81, 2205–2268.

DiNardo, J., N. Fortin, and T. Lemieux (1996): "Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach," *Econometrica*, 64, 1001–1044.

Donald, S., D. Green, and H. Paarsch (2000): "Differences in wage distributions between Canada and the United States: An application of a flexible estimator of distribution functions in the presence of covariates," *Review of Economic Studies*, 67, 609–633.

Firpo, S., N. Fortin, and T. Lemieux (2007): "Decomposing wage distributions using recentered influence function regressions," *NBER Working Paper*.

——— (2009): "Unconditional quantile regressions," *Econometrica*, 77, 953–973.

——— (2013): "Occupational tasks and changes in the wage structure," *Working Paper*.

Foresi, S. and F. Peracchi (1995): "The Conditional Distribution of Excess Returns: An Empirical Analysis," *Journal of the American Statistical Association*, 90, 451–466.

Fortin, N., T. Lemieux, and S. Firpo (2011): "Decomposition Methods in Economics," *Handbook of Labor Economics*, 4, 1–102.

Genest, C., K. Ghoudi, and L. Rivest (1995): "A semiparametric estimation procedure of dependence parameters in multivariate families of distributions," *Biometrika*, 82, 543–552.

Gosling, A., S. Machin, and C. Meghir (2000): "The Changing Distribution of Male Wages in the UK," *Review of Economic Studies*, 67, 635–666.

Joe, H. (1997): *Multivariate models and dependence concepts*, Chapman & Hall/CRC.

Koenker, R. (2005): *Quantile regression*, Cambridge University Press.

Koenker, R. and G. Bassett (1978): "Regression quantiles," *Econometrica*, 46, 33–50.

Lemieux, T. (2006): "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?" *American Economic Review*, 96, 461–498.

——— (2008): "The changing nature of wage inequality," *Journal of Population Economics*, 21, 21–48.

Machado, J. and J. Mata (2005): "Counterfactual decomposition of changes in wage distributions using quantile regression," *Journal of Applied Econometrics*, 20, 445–465.

Melly, B. (2005): "Decomposition of differences in distribution using quantile regression," *Labour Economics*, 12, 577–590.

Nelsen, R. (2006): *An introduction to copulas*, Springer Verlag.

Newey, W. and D. McFadden (1994): "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 4, 2111–2245.

Nikoloulopoulos, A. and D. Karlis (2009): "Finite normal mixture copulas for multivariate discrete data modeling," *Journal of Statistical Planning and Inference*, 139, 3878–3890.

Oaxaca, R. (1973): "Male-female wage differentials in urban labor markets," *International Economic Review*, 14, 693–709.

Panagiotelis, A., C. Czado, and H. Joe (2012): "Pair Copula Constructions for Multivariate Discrete Data," *Journal of the American Statistical Association*, 107, 1063–1072.

Rothe, C. (2010): "Nonparametric Estimation of Distributional Policy Effects," *Journal of Econometrics*, 155, 56–70.

——— (2012): "Partial Distributional Policy Effects," *Econometrica*, 80, 2269–2301.

Rothe, C. and D. Wied (2013): "Misspecification testing in a class of conditional distributional models," *Journal of the American Statistical Association*, 108, 314–324.

Sklar, A. (1959): "Fonctions de répartition à n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, 8, 229–231.

SMITH, M. AND M. KHALED (2012): "Estimation of copula models with discrete margins via Bayesian data augmentation," *Journal of the American Statistical Association*, 107, 290–303.

TRIVEDI, P. AND D. ZIMMER (2007): *Copula modeling: an introduction for practitioners*, Now Publishers Inc.

VAN DER VAART, A. (2000): *Asymptotic Statistics*, Cambridge University Press.

Table 1: Descriptive Statistics

	1983–1985					2003–2005				
	Mean	SD	Q10	Q50	Q90	Mean	SD	Q10	Q50	Q90
Log Wage $\times 100$	178.45	52.43	104.98	181.88	245.73	184.85	58.28	110.79	182.67	264.84
Part-time status	0.09	–	–	–	–	0.09	–	–	–	–
Nonwhite	0.11	–	–	–	–	0.13	–	–	–	–
Union coverage	0.27	–	–	–	–	0.15	–	–	–	–
Married	0.67	–	–	–	–	0.62	–	–	–	–
Education	12.87	2.91	10.00	12.00	17.00	13.42	2.79	11.00	13.00	18.00
Experience	17.20	12.31	3.00	14.00	36.00	19.47	11.49	4.00	19.00	35.00

Table 2: Estimated Decomposition of Differences in Distribution of Log Hourly Wages ($\times 100$) of Workers in 2003–2005 and 1983–1985 using CPS Data.

	Mean		Q90		Q50		Q10	
Total Difference Δ_O^ν	6.479	(0.189)	18.756	(0.327)	0.862	(0.289)	5.401	(0.267)
Structure Effect Δ_S^ν	1.007	(0.167)	8.398	(0.338)	-5.632	(0.308)	3.660	(0.272)
Composition Effect Δ_X^ν	5.472	(0.127)	10.358	(0.190)	6.493	(0.217)	1.742	(0.183)
Dependence Effect Δ_D^ν	-0.382	(0.033)	0.737	(0.063)	-0.448	(0.063)	-1.647	(0.096)
Marginal Distr. Effect Δ_M^ν	5.854	(0.121)	9.621	(0.171)	6.941	(0.211)	3.388	(0.181)
"Direct" Contribution to Composition Effect ($\Delta_M^\nu(\mathbf{e}^j)$)								
Part-time	-0.051	(0.013)	0.016	(0.004)	-0.045	(0.012)	-0.143	(0.038)
Nonwhite	-0.263	(0.018)	-0.191	(0.016)	-0.329	(0.026)	-0.265	(0.021)
Union coverage	-2.327	(0.035)	1.092	(0.058)	-3.913	(0.085)	-2.636	(0.093)
Married	-0.526	(0.019)	-0.254	(0.019)	-0.634	(0.035)	-0.640	(0.034)
Education	4.349	(0.074)	4.112	(0.104)	5.631	(0.138)	4.005	(0.126)
Experience	4.307	(0.058)	4.468	(0.100)	5.629	(0.121)	3.647	(0.112)
"Two-Way" Interactions ($\Delta_M^\nu(\mathbf{k})$ with $\lvert\mathbf{k}\rvert = 2$)								
Part-time:Nonwhite	0.000	(0.000)	-0.001	(0.000)	-0.001	(0.003)	0.001	(0.005)
Part-time:Union coverage	-0.007	(0.002)	-0.005	(0.002)	-0.007	(0.004)	-0.019	(0.011)
Part-time:Married	-0.001	(0.000)	-0.001	(0.000)	-0.001	(0.004)	-0.011	(0.008)
Part-time:Education	-0.001	(0.000)	0.001	(0.002)	-0.003	(0.003)	-0.037	(0.019)
Part-time:Experience	0.003	(0.001)	0.003	(0.003)	0.001	(0.004)	-0.028	(0.018)
Nonwhite:Union coverage	-0.014	(0.002)	-0.004	(0.006)	-0.001	(0.018)	-0.038	(0.018)
Nonwhite:Married	0.003	(0.000)	-0.001	(0.002)	0.001	(0.012)	-0.024	(0.013)
Nonwhite:Education	0.004	(0.001)	0.018	(0.008)	-0.005	(0.018)	-0.069	(0.020)
Nonwhite:Experience	-0.015	(0.001)	0.009	(0.008)	-0.046	(0.017)	-0.080	(0.020)
Union coverage:Married	-0.035	(0.003)	-0.022	(0.006)	-0.030	(0.032)	-0.070	(0.046)
Union coverage:Education	0.315	(0.009)	0.241	(0.058)	0.265	(0.122)	-0.245	(0.105)
Union coverage:Experience	0.084	(0.007)	0.165	(0.053)	-0.158	(0.069)	-0.303	(0.108)
Married:Education	0.002	(0.003)	0.013	(0.015)	0.027	(0.022)	-0.169	(0.043)
Married:Experience	-0.013	(0.004)	0.013	(0.012)	-0.001	(0.032)	-0.188	(0.041)
Education:Experience	0.016	(0.006)	-0.094	(0.102)	-0.966	(0.146)	0.911	(0.138)

Note: Bootstrapped standard errors (200 replications) are in parenthesis.

Table 3: Estimated Decomposition of Differences in Distribution of Log Hourly Wages ($\times 100$) of Workers in 2003–2005 and 1983–1985 using CPS Data.

	Std. Dev		Q90-10		Q90-50		Q50-10	
Total Difference Δ_O^ν	5.984	(0.118)	13.355	(0.387)	17.895	(0.352)	-4.540	(0.316)
Structure Effect Δ_S^ν	3.456	(0.126)	4.738	(0.421)	14.030	(0.392)	-9.291	(0.373)
Composition Effect Δ_X^ν	2.528	(0.062)	8.616	(0.234)	3.865	(0.237)	4.752	(0.192)
Dependence Effect Δ_D^ν	0.681	(0.030)	2.384	(0.119)	1.185	(0.074)	1.199	(0.105)
Marginal Distr. Effect Δ_M^ν	1.847	(0.050)	6.232	(0.217)	2.680	(0.226)	3.553	(0.194)
"Direct" Contribution to Composition Effect ($\Delta_M^\nu(\mathbf{e}^j)$)								
Part-time	0.052	(0.014)	0.159	(0.042)	0.061	(0.016)	0.099	(0.026)
Nonwhite	0.037	(0.004)	0.073	(0.015)	0.137	(0.018)	-0.064	(0.018)
Union coverage	1.153	(0.021)	3.728	(0.112)	5.006	(0.105)	-1.278	(0.105)
Married	0.184	(0.009)	0.386	(0.034)	0.380	(0.036)	0.006	(0.035)
Education	-0.050	(0.023)	-0.107	(0.138)	-1.518	(0.134)	1.625	(0.147)
Experience	0.156	(0.026)	0.821	(0.136)	-1.162	(0.141)	1.982	(0.146)
"Two-Way" Interactions ($\Delta_M^\nu(\mathbf{k})$ with $\lvert\mathbf{k}\rvert = 2$)								
Part-time:Nonwhite	-0.001	(0.000)	-0.002	(0.001)	-0.001	(0.003)	-0.002	(0.006)
Part-time:Union coverage	-0.001	(0.001)	0.012	(0.011)	0.001	(0.002)	0.011	(0.011)
Part-time:Married	0.000	(0.000)	-0.001	(0.003)	0.001	(0.003)	-0.003	(0.005)
Part-time:Education	0.003	(0.001)	0.043	(0.015)	-0.003	(0.003)	0.046	(0.016)
Part-time:Experience	0.004	(0.001)	0.036	(0.013)	-0.001	(0.003)	0.037	(0.014)
Nonwhite:Union coverage	-0.002	(0.001)	0.040	(0.019)	0.005	(0.020)	0.035	(0.025)
Nonwhite:Married	-0.002	(0.001)	-0.003	(0.010)	-0.026	(0.021)	0.023	(0.024)
Nonwhite:Education	0.005	(0.001)	0.092	(0.019)	0.026	(0.018)	0.065	(0.024)
Nonwhite:Experience	0.009	(0.001)	0.097	(0.020)	0.052	(0.020)	0.045	(0.024)
Union coverage:Married	-0.007	(0.002)	0.057	(0.047)	0.014	(0.042)	0.042	(0.055)
Union coverage:Education	0.142	(0.005)	0.486	(0.121)	-0.024	(0.139)	0.510	(0.156)
Union coverage:Experience	0.174	(0.005)	0.514	(0.114)	0.312	(0.139)	0.202	(0.153)
Married:Education	0.013	(0.002)	0.185	(0.041)	-0.026	(0.034)	0.211	(0.052)
Married:Experience	0.022	(0.003)	0.201	(0.043)	0.011	(0.035)	0.189	(0.053)
Education:Experience	-0.036	(0.005)	-1.005	(0.170)	0.872	(0.178)	-1.877	(0.148)

Note: Bootstrapped standard errors (200 replications) are in parenthesis.

Table 4: Estimated Copula Parameters.

	Nonwhite		Union		Married		Education		Experience	
	1985	2005	1985	2005	1985	2005	1985	2005	1985	2005
Part-time	0.092	0.049	-0.113	-0.095	-0.412	-0.392	-0.063	-0.165	-0.314	-0.293
Nonwhite			0.089	0.043	-0.144	-0.120	-0.094	0.028	0.008	-0.056
Union					0.180	0.119	-0.170	-0.008	0.260	0.195
Married							0.049	0.199	0.580	0.457
Education									-0.168	-0.041

Estimates of the parameters of the Gaussian copula that determine the pairwise dependence structure between the two respective covariates.